Amino Acid Sequence of the N-Terminal Forty-Two Amino Acid Residues of the C Chain of Subcomponent C1q of the First Component of Human Complement

By KENNETH B. M. REID

Medical Council Immunochimistry Unit, Department of Biochemistry,
University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

(Received 13 August 1976)

1. The sequence of the N-terminal 42 amino acid residues and the identity of residue 45 of the C chain of subcomponent C1q were established by the use of the automatic protein sequencer. 2. A comparison of the amino acid sequences of the A and C chains of subcomponent C1q shows that they are identical at 18 positions out of the first 45. However, 12 of the amino acid residues in the positions of identity are glycine residues occurring in the repeating triplet sequence Gly-X-Y. 3. Position 36 in the C chain was found to be alanine, which was unexpected since the residue in this position would have to be glycine if the repeating triplet sequence Gly-X-Y were to extend, uniformly, throughout the entire length of the C-chain collagen-like region. This break in the repeating triplet sequence would prevent residue 36 in the C chain from taking part in collagen-like triple-helix formation. 4. The sequence information presented here therefore indicates that there should be a break, or distortion, located approximately half-way along each of the six collagen-like triple-helical regions proposed to be present in subcomponent C1q [Reid & Porter (1976) *Biochem. J.* 155, 19–23] and this is consistent with what is seen in electron micrographs of intact and pepsin-digested subcomponent C1q.

Subcomponent C1q is composed of two types of non-covalently linked subunits each having mol.wt. 49000–50000. One subunit is composed of an A chain and a B chain, linked by a single disulphide bond, and the second is composed of two C chains, also linked by a single disulphide bond (Yonemasu & Stroud, 1972; Reid & Porter, 1976). Reduction, or performic acid oxidation, of the disulphide bonds in subcomponent C1q yields six A, six B and six C chains each approx. 200 amino acid residues long and of mol.wt. 23000–24000 (Reid *et al.*, 1972). The six A–B dimer subunits and three C–C dimer subunits combine to yield the intact molecule of mol.wt. 410000. Structural studies have established that there is a collagen-like region of 78 amino acid residues located near the N-terminus of each of the three chains of subcomponent C1q (Reid, 1974, 1976). It has been proposed that the collagen-like regions from each of the three chains will form a triple-helical structure throughout their length (Reid & Porter, 1976). Physical and electron-microscopy studies have yielded results consistent with there being a collagen-like triple-helical structure in subcomponent C1q (Brodsky-Doyle *et al.*, 1976). The N-terminal sequence of over half the collagen-like region of the C chain, which is presented in the present paper, shows that the proposal on the formation of a triple-helical structure in subcomponent C1q requires modification to allow for a break in the re-

peating nature of the collagen-like sequence in one of the chains.

Materials and Methods

Automated sequence determination, quantitative amino acid analysis, manual dansyl-Edman degradation, electrophoresis and chromatography

The procedures were performed as described previously (Reid, 1974, 1976).

Isolation of peptide P1-CNBrIII-CMCV

P1, the pepsin-derived fragment of subcomponent C1q, which contains all the collagen-like features present in the intact molecule, was isolated as described by Reid (1976). Peptide P1 (42 mg) was dissolved in aq. 70% (v/v) formic acid (7.0 ml) containing CNBr (200 mg), and the mixture incubated, under N₂, for 16 h at 25°C; water (150 ml) was then added and the digest freeze-dried. The freeze-dried material (42 mg) was suspended in 0.1 M-acetic acid (10 ml) and the supernatant of the spun suspension applied to a column (2.5 cm × 80 cm) of Sephadex G-150 which had been equilibrated with 0.1 M-acetic acid. The fraction P1-CNBrIII, which was eluted between 396 ml and 431 ml, was pooled and freeze-dried. The freeze-dried fraction P1-CNBrIII (9 mg) was dissolved in 10 mM-sodium citrate/citric acid buffer, pH 3.6, and applied to a CM-cellulose column (2 cm × 15 cm)

equilibrated with the same buffer. A linear gradient of NaCl (0–0.2M) in the starting buffer and in a total volume of 400ml was applied to the column. The column effluent was monitored at 230nm and the peak corresponding to peptide P1-CNBrIII-CMCV was eluted after 150ml of the gradient had been run. Peptide P1-CNBrIII-CMCV was de-salted on a column (4 cm × 30 cm) of Sephadex G-25, equilibrated with 0.1 M-acetic acid, and then freeze-dried.

Trypsin digestion of peptide P1-CNBrIII-CMCV

Peptide P1-CNBrIII-CMCV (150nmol) was dissolved in 200mM-NH₄HCO₃ buffer (200μl), pH 8.1, and digested with trypsin (70μg) for 2h at 37°C. The digest was fractionated by electrophoreses at pH 3.5. Peptides T1 and T2 had mobilities of +0.50 and +0.26 respectively (relative to lysine = +1.0) and were found to correspond to residues 1–9 and 10–28 of peptide P1-CNBrIII-CMCV.

Isolation of fragment P1-CMC2

Fragment P1-CMC2, which was known to be a dimer of the first 94 amino acid residues of the C chain, was isolated, reduced and alkylated as described by Reid (1967).

Results

N-Terminal sequence of fragment P1-CMC2

It was known from previous results that the reduced and alkylated fragment P1-CMC2 was composed of the N-terminal 94 amino acid residues of the C chain (Reid, 1976), and the sequence of residues 1–10 of this fragment has been published (Reid, 1976). Further automated sequential Edman degradation, on a larger amount of the reduced and alkylated fragment P1-CMC2, established that the sequence of residues 11–19 of fragment P1-CMC2, and therefore of the C chain, was Hyp-Gly-Leu-Hyp-Gly-Ala-Hyp-Gly-Lys (Fig. 1). Comparison of this sequence with the sequence of the first nine amino acid residues of peptide P1-CNBrIII-CMCV established that peptide P1-CNBrIII-CMCV represents the positions 11–67 of the C chain (Fig. 1). The experimental data proving the identity of residues 11–19 of fragment P1-CMC2 are not given here, since the proof of this sequence is given below in the studies on the N-terminal sequence of peptide P1-CNBrIII-CMCV.

N-Terminal sequence of peptide P1-CNBrIII-CMCV

The amino acid composition of peptide P1-CNBrIII-CMCV is shown in Table 1, and the proposed amino acid sequence of the first 32 residues of this peptide is given in Table 2 and Fig. 1. The yields of hydroxyproline obtained, on back hydrolysis of the phenylthiohydantoin in HI, at steps 1, 4, 7, 16, 18 and 25 were low compared with the yields of glycine and leucine at neighbouring positions in the sequencer run (Table 2). For example, the yield of hydroxypro-

Table 1. *Amino acid compositions of the peptide P1-CNBrIII-CMCV and two of its tryptic peptides*
Values are residues/mol of peptide and values less than 0.1 are omitted. Full experimental details are given in the text.

Amino acid	Amino acid composition (mol of residues/mol)		
	P1-CNBrIII-CMCV	T1	T2
Cys(O ₃ H)	—	—	—
Hyp	9.7 (10)	2.6 (3)*	3.4 (3)
Asp	2.9 (3)	—	1.6 (2)
Met(O ₂)	—	—	—
Thr	—	—	—
Ser	—	—	—
Hse	0.5 (1)	—	—
Glu	3.2 (3)	—	1.2 (1)
Pro	4.0 (4)	0.5 (0)*	1.6 (2)
Gly	17.9 (18)	3.3 (3)	5.2 (5)
Ala	1.9 (2)	0.8 (1)	0.9 (1)
Val	—	—	—
Ile	3.0 (3)	—	2.0 (2)
Leu	2.6 (3)	0.8 (1)	1.2 (1)
Tyr	1.0 (1)	—	0.6 (1)
Phe	—	—	—
His	0.9 (1)	—	—
Hyl	3.5 (4)	0.3 (0)	0.5 (1)
Lys	2.0 (2)	0.8 (1)	—
Arg	1.0 (1)	—	0.2 (0)
Total residues	56	9	19
Yield (%)	25	60	67
N-Terminal	Hyp	Hyp	Asx

* There may be partial hydroxylation of one or more of the proline residues at positions 1, 4 and 7 in peptide T1. Therefore the total proline plus hydroxyproline content of this peptide has been taken as 3.0.

line at step 1 was 12nmol, whereas 42nmol of glycine and 48nmol of leucine were found at positions 2 and 3 respectively. However, at each step, only one major residue, when identified, was seen on g.l.c., t.l.c. and amino acid analysis after back-hydrolysis. This low recovery of hydroxyproline could be due to difficulty in extraction of the phenylthiohydantoin derivative or to excessive losses during hydrolysis in HI. Carry-over of an amino acid, identified at one step, to the next step did not exceed 18%, except for glycine at steps 29–30 and 32–33, where the carry-over was approx. 30%. The yield of alanine at step 26 was 2.8 times that of the glycine background. Residue 28 was not identified from the sequencer run, but was positioned by the isolation of the tryptic peptides T1 and T2 as described below.

Peptides T1 and T2

The amino acid compositions and N-terminal amino acids of peptides T1 and T2 are given in Table 1. Peptide T1 corresponds to residues 1–9 of peptide

Table 2. *Proposed sequence of the first 32 amino acid residues of the peptide P1-CNBrIII-CMCV derived by CNBr digestion of the pepsin fragment of subcomponent C1q*

Peptide P1-CNBrIII-CMCV (140nmol) was subjected to 40 cycles of Edman degradation in the Beckman 890C protein sequencer by using the 0.1M-Quadrol protein programme. Full experimental details are given in Reid (1976). The amino acid phenylthiohydantoin derivatives were identified by g.l.c., t.l.c. and back-hydrolysis in HI followed by amino acid analysis. Yields of amino acid phenylthiohydantoin derivatives were calculated from the g.l.c. peak areas and amino acid analysis after back-hydrolysis. X represents an unidentified residue.

Residue no.	Method of identification of amino acid phenylthiohydantoin derivative			Proposed sequence	Yield of amino acid phenylthiohydantoin derivative (nmol)
	G.l.c.	T.l.c.	Amino acid analysis		
1	—	Hyp	Hyp	Hyp	12
2	Gly	Gly	Gly	Gly	42
3	Ile/Leu	Ile/Leu	Leu	Leu	48
4	—	Hyp	Hyp	Hyp	6
5	Gly	Gly	Gly	Gly	30
6	Ala	Ala	Ala	Ala	30
7	—	Hyp	Hyp	Hyp	4
8	Gly	Gly	Gly	Gly	26
9	—	Lys	Lys	Lys	10
10	—	Asp	Asx	Asp	10
11	Gly	Gly	Gly	Gly	23
12	—	Tyr	Tyr	Tyr	8
13	—	Asp	Asx	Asp	7
14	Gly	Gly	Gly	Gly	19
15	Ile/Leu	Ile/Leu	Leu	Leu	20
16	—	—	Hyp	Hyp	2
17	Gly	Gly	Gly	Gly	16
18	Pro	Pro	Pro	Pro	8
19	—	Hyp	Hyp	Hyp	—
20	Gly	Gly	Gly	Gly	7
21	—	Glu	Glx	Glu	6
22	Pro	Pro	Pro	Pro	3
23	Gly	Gly	Gly	Gly	9
24	Ile/Leu	Ile/Leu	Ile	Ile	6
25	—	—	Hyp	Hyp	—
26	Ala	Ala	Ala	Ala	6
27	Ile/Leu	Ile/Leu	Ile	Ile	6
28	—	—	—	(Hyl)*	—
29	—	Gly	Gly	Gly	5
30	Ile/Leu	Ile/Leu	Ile	Ile	5
31	—	Asp	Asx	Asp	—
32	—	Gly	Gly	Gly	—
33	—	—	—	X	—
34	—	—	—	X	—
35	—	Gly	Gly	Gly	—

* This residue was identified as hydroxylysine by isolation of the tryptic peptides of P1-CNBrIII-CMCV.

P1-CNBrIII-CMCV and peptide T2 corresponds to residues 10–28 (Fig. 1). Since there is only one basic amino acid residue in peptide T2, i.e. hydroxylysine, this amino acid was taken to be the C-terminal amino acid of peptide T2 (Fig. 1).

Discussion

The N-terminal amino acid sequences of the A and C chains of subcomponent C1q were found to be identical at 18 out of the first 45 positions (Fig. 1).

However, 12 of these positions of identity are occupied by glycine in the repeating triplet sequence Gly-X-Y. The other six positions of identity are at Cys-4, Hyp-17, Leu-25, Glu-31, Hyp-35 and Ile-37 (Fig. 1). The location of a half-cystine residue at position 4 in both chains may be of importance in alignment of the chains for triple-helix formation, since the half-cystine at A4 is linked to a half-cystine in the sequence B2–B6 and the half-cystine at C4 is linked to another C4 half-cystine (Reid, 1976). A comparison of the chains made by aligning the glycine at A9 with the

glycine at C2 (i.e. the first glycine residue of the collagen-like region of each chain) decreases the number of non-glycine identities to four, but interestingly lines up all four hydroxyproline residues in the first 45 positions in the A chain with those in the C chain (Fig. 1).

One unusual feature of the C chain *N*-terminal amino acid sequence is the presence of an alanine residue at position 36 in the C chain, where a glycine residue might have been expected, since this involves a break in the Gly-X-Y repeating triplet sequence. Studies, involving the use of the automatic sequencer and collagenase digestion, indicate that amino acid residues 45–80 of the C chain are composed of Gly-X-Y repeating triplet sequence, although a number of amino acid residues in the X and Y positions have not yet been unequivocally identified (K. B. M. Reid, unpublished work). The location of the alanine residue at position 36 in the C chain shows that a modification must be made to the proposal that the A, B and C chains may form a triple-helical region by an interaction throughout the entire length of their collagen-like amino acid sequences of 78 residues (Reid & Porter, 1976). This is because of the requirement for glycine in every third position in order to form a continuous collagen-type molecule in which the glycine residues are packed at the centre of the cylindrical triple-helical structure. The glycine residues are packed in such a way that there is no room for an L-amino acid side chain (Traub & Piez, 1971). It is possible to form a triple-helical structure with a D-alanine residue in the place of a glycine residue, but it seems unlikely that D-alanine would be present in the C chain of subcomponent C1q. Therefore if the three chains of subcomponent C1q do form a triple helix through the association of their collagen-like regions, it would be expected that there would be a break in the triple-helical structure at the point involving the alanine at position 36 in the C chain. The amino acid-sequence studies thus predict that there should be a distortion in the subcomponent C1q

molecule approx. $36 \times 0.291 \text{ nm} = 10.5 \text{ nm}$ ($36 \times 2.91 \text{ \AA} = 105 \text{ \AA}$) [0.291 nm ($2.91 \text{ \AA}$) being taken as the length of the cross-section of a one-residue-long stretch of a collagen triple helix] from the *N*-terminus of each of the six collagen-like triple helices proposed to be in subcomponent C1q (Reid & Porter, 1976). Consistent with this prediction is the observation that in electron micrographs of subcomponent C1q the molecule does appear to 'bend' approximately half-way up [11.2 nm ($112 \text{ \AA}$)] the proposed collagen-like regions, i.e. at the point where the central fibril-like region spreads out to give the six connecting strands (Shelton *et al.*, 1972; Knobel *et al.*, 1975). These 'bends' in the molecule are still seen in the electron microscope after removal, by limited pepsin digestion, of the six peripheral globular regions which are joined to the six connecting strands (Brodsky-Doyle *et al.*, 1976). This observation suggests that the 'bends' observed in the proposed collagen-like regions are not induced by the steric repulsion of the six peripheral globular regions, but are caused by the break in the helical structure necessitated by the presence of an alanine residue at position 36 in the C chain.

References

- Brodsky-Doyle, B., Leonard, K. R. & Reid, K. B. M. (1976) *Biochem. J.* **159**, 279–286
Knobel, H. R., Villiger, W. & Isliker, H. (1975) *Eur. J. Immunol.* **5**, 78–82
Reid, K. B. M. (1974) *Biochem. J.* **141**, 189–203
Reid, K. B. M. (1976) *Biochem. J.* **155**, 5–17
Reid, K. B. M. & Porter, R. R. (1976) *Biochem. J.* **155**, 19–23
Reid, K. B. M., Lowe, D. M. & Porter, R. R. (1972) *Biochem. J.* **130**, 749–763
Shelton, E., Yonemasu, K. & Stroud, R. M. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 65–68
Traub, W. & Piez, K. A. (1971) *Adv. Protein Chem.* **25**, 243–352
Yonemasu, K. & Stroud, R. M. (1972) *Immunochemistry* **9**, 545–554